ENGG 2430 / ESTR 2004: Probability and Statistics

Spring 2019

4. Expectation and Variance
Joint PMFs

Andrej Bogdanov

Expected value

The expected value (expectation) of a random variable X with p.m.f. p is

$1, 2, 3 \ldots$

$$E[X] = \sum_x x\, p(x)$$

$$= 1 \cdot p(1) + 2p(2) + \ldots$$

Example

$N =$ **number of** H**s**

x	0	1
$p(x)$	$1/2$	$1/2$

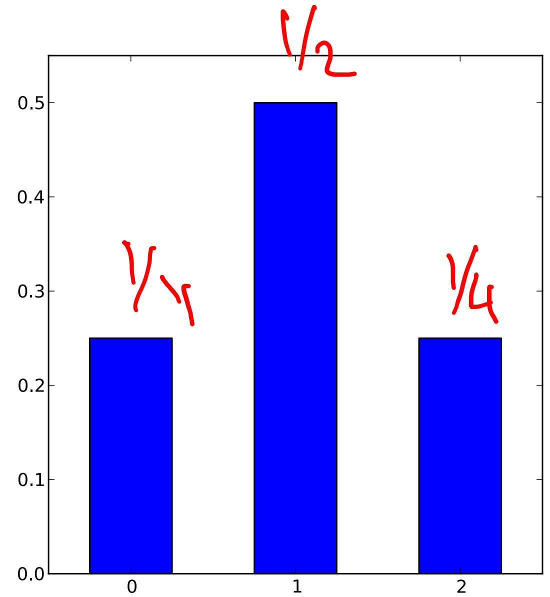
$$E[N] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

Expected value

Example

$N =$ **number of** H**s**

$$E[N] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4}$$
$$= 1$$



The expectation is the average value the random variable takes when experiment is done many times

$F =$ face value of fair 6-sided die

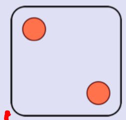
x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E[F] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6}$$

$$= (1 + 2 + \ldots + 6) \cdot \frac{1}{6}$$

$$= \frac{21}{6} = \frac{7}{2} = 3.5$$

P = PROFIT

If  appears k times, you **win** \$$k$. Binomial$(3, \frac{1}{6})$

If it **doesn't appear**, you lose \$1.

x	-1	1	2	3
$p(x)$	$\left(\frac{5}{6}\right)^3$	$3 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2$	$3 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6}$	$\left(\frac{1}{6}\right)^3$

$$E[P] = -1 \cdot \left(\frac{5}{6}\right)^3 + 1 \cdot 3 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 + 2 \cdot 3\left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} + 3 \cdot \left(\frac{1}{6}\right)^3$$

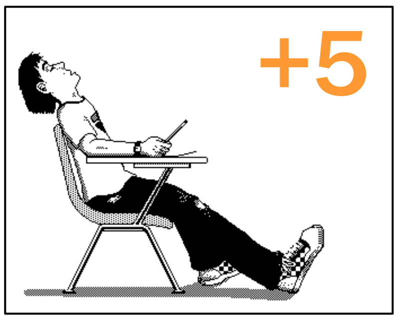
$$= -0.03\ldots$$

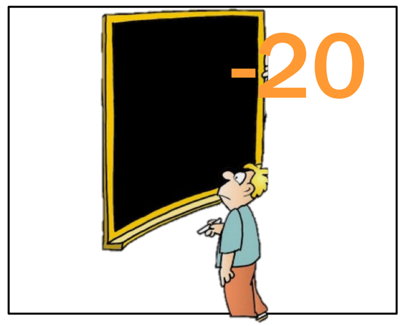
Utility

Should I go to tutorial?

not called | called

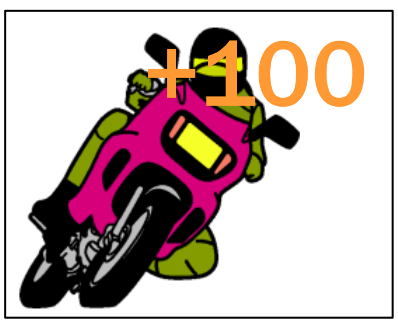
*C*ome

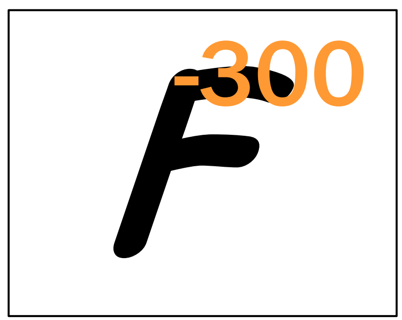




$$E[C] = 5 \cdot \frac{35}{40} - 20 \cdot \frac{5}{40} = 1.9$$

*S*kip





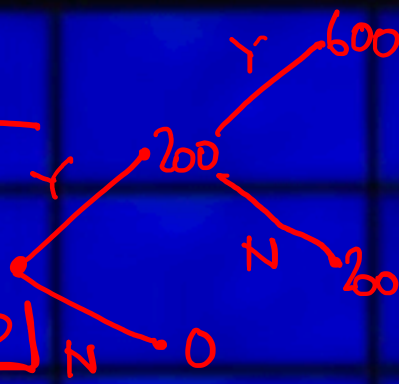
$$E[S] = 100 \cdot \frac{35}{40} - 300 \cdot \frac{5}{40} = 50$$

35/40 | 5/40

A : FIRST TAKE $200

x	0	200	600
p(x)	20% =0.2	80%·50% =0.4	80%·50% =0.4

$E[A] = 200 \cdot 0.4 + 600 \cdot 0.4 = \boxed{320}$



B : FIRST TAKE $400

x	0	400	600
p(x)	50%	50%·20% =0.1	50%·80% =0.4

$E[B] = 400 \cdot 0.1 + 600 \cdot 0.4 = \boxed{280}$

VIDEO GAMES	
$200	80%
$400	50%
~~$600~~	INDEP
~~$800~~	
~~$1000~~	

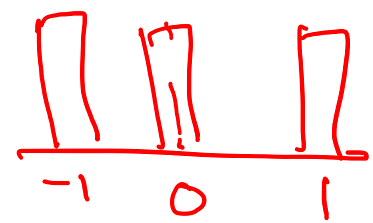
Expectation of a function

p.m.f. of X:

x	0	1	2
$p(x)$	$1/3$	$1/3$	$1/3$

$\mathbf{E}[X] =$ $0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$

x	-1	0	1
p(x)	1/3	1/3	1/3



$\mathbf{E}[X-1] =$ $-1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$

$\mathbf{E}[(X-1)^2] =$ $\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$

x	0	1
p(x)	1/3	2/3

Expectation of a function, again

p.m.f. of X:

x	0	1	2
$p(x)$	$1/3$	$1/3$	$1/3$

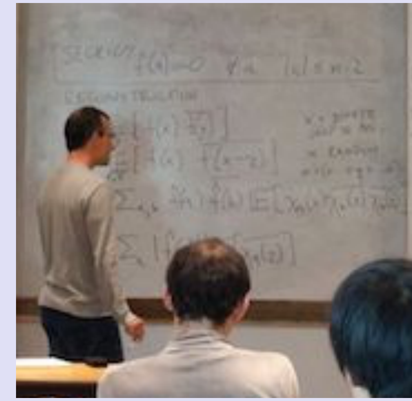
$\mathbf{E}[X] = 1$

$\mathbf{E}[X-1] = \frac{1}{3}\cdot(-1) + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1 = 0$

$\mathbf{E}[(X-1)^2] = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1 = \frac{2}{3}$

$$\boxed{\mathbf{E}[f(X)] = \sum_x f(x)\, p(x)}$$

1km



60%

5km/h

40%

30km/h

AVERAGE TIME?

$$\text{speed } V = \frac{\text{distance}}{\text{time } T} \qquad T = \frac{1}{V}$$

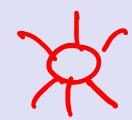
$$E[V] = 60\% \cdot 5 + 40\% \cdot 30 = 15$$

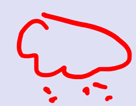
$$\frac{1}{E[V]} = 0.6 \text{ hrs}$$

$$E[T] = 60\% \cdot \frac{1}{5} + 40\% \cdot \frac{1}{30}$$

$$\approx 0.133 \text{ hrs}$$

x	$1/5$	$1/30$
$p(x)$	60%	40%





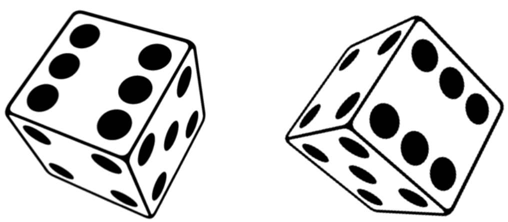
$$E[T] = E\left[\frac{1}{V}\right] \neq \frac{1}{E[V]}$$

Joint probability mass function

The joint PMF of random variables X, Y is the bivariate function

$$p(x, y) = \mathbf{P}(X = x, Y = y)$$

Example



X Y

$p(x,y) = \frac{1}{36}$

FOR ALL x, y

x \ y	1	2	3	4	5	6
1	1/36	1/36		⋮		
2				⋮		
3	…	…	…	1/36	…	…
4				⋮		
5				⋮		
6				⋮		

There is a bag with 4 cards:

1 2 3 4

X = 1ST CARD

Y = 2ND CARD

You draw two without replacement. What is the joint PMF of the face values?

$X \backslash Y$	1	2	3	4
1	0	1/12	1/12	1/12
2	1/12	0	1/12	1/12
3	1/12	1/12	0	1/12
4	1/12	1/12	1/12	0

$x \neq y$

$\downarrow$

$$P(X=x, Y=y)$$

$$= P(X=x)\,P(Y=y \mid X=x)$$

$$= \frac{1}{4} \cdot \frac{1}{3}$$

What is the PMF of the sum? S

$X \backslash Y$	1	2	3	4
1	0	1/12	1/12	1/12
2	1/12	0	1/12	1/12
3	1/12	1/12	0	1/12
4	1/12	1/12	1/12	0

s	2	3	4	5	6	7	8
$p(s)$	0	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{6}{12}$	$\frac{2}{12}$	0

What is the expected value?

$$E[S] = 3 \cdot \frac{2}{12} + 4 \cdot \frac{2}{12} + 5 \cdot \frac{4}{12} + 6 \cdot \frac{2}{12} + 7 \cdot \frac{2}{12} = 5$$

PMF and expectation of a function

Z

$Z = f(X, Y)$ **has PMF**

$$p_Z(z) = \sum_{x, y: f(x, y) = z} p_{XY}(x, y)$$

and expected value

$$\mathbf{E}[Z] = \sum_{x, y} f(x, y)\, p_{XY}(x, y)$$

What if the cards are drawn with replacement?

JOINT PMF: $p(x,y) = \frac{1}{16}$ FOR ALL x, y

$$E[S] = \frac{1}{16}(2+3+4+5 + 3+4+5+6 + 4+5+6+7 + 5+6+7+8)$$

$$= 5$$

y \ x	1	2	3	4
1	1/16	1/16	1/16	1/16
2	1/16	1/16	1/16	1/16
3	1/16	1/16	1/16	1/16
4	1/16	1/16	1/16	1/16

Marginal probabilities

$Y \backslash X$	1	2	3	4	
1	0	1/12	1/12	1/12	1/4
2	1/12	0	1/12	1/12	1/4
3	1/12	1/12	0	1/12	1/4
4	1/12	1/12	1/12	0	1/4
	1/4	1/4	1/4	1/4	1

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Linearity of expectation

$$E[X+Y] = \sum_{x,y} (x+y) \cdot p_{XY}(x,y)$$
$$= \sum x\, p_{XY}(x,y) + \sum y\, p_{XY}(x,y)$$
$$= \sum x\, p_X(x) + \sum y\, p_Y(y)$$
$$= E[X] + E[Y]$$

For every two random variables X and Y

$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

1	2	3	4

WITHOUT REPLACEMENT

x	1	2	3	4
$p_X(x)$	1/4	1/4	1/4	1/4

y	1	2	3	4
$p_Y(y)$	1/4	1/4	1/4	1/4

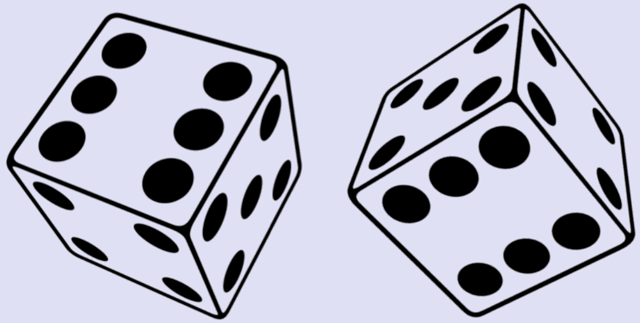
$$\left.\begin{aligned} E[X] &= 2.5 \\ E[Y] &= 2.5 \end{aligned}\right\} E[X+Y] = 5$$

WITH REPLACEMENT

SAME

SAME

SAME



$$\mathbf{E}[X+Y] = ?$$

1ST DIE

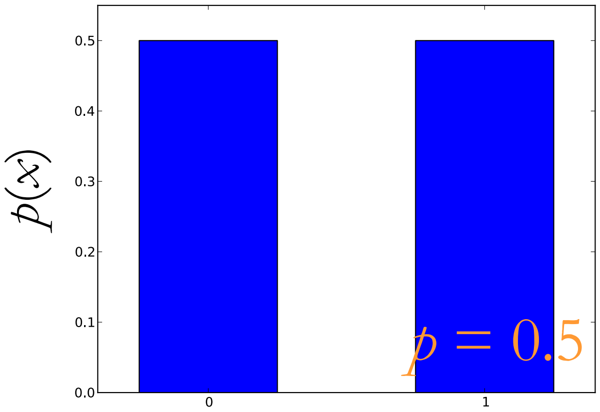
2ND DIE

$$E[X+Y] = E[X] + E[Y]$$
$$= 3.5 + 3.5$$
$$= 7$$

The indicator (Bernoulli) random variable

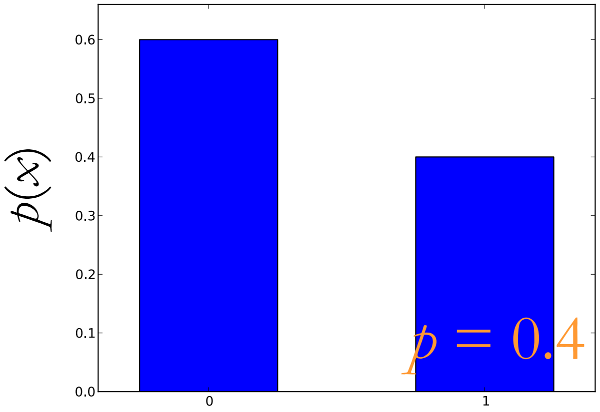
Perform a trial that succeeds with probability p and fails with probability $1-p$.

x	0	1
$p(x)$	$1-p$	p



If X is $\mathrm{Bernoulli}(p)$ then

$$E[X] = p$$



Mean of the Binomial

$\text{Binomial}(n, p)$: **Perform** n **independent trials, each of which succeeds with probability** p.

X = **number of successes** $= X_1 + X_2 + \cdots + X_n$

$X_1, X_2, \ldots, X_n$ ARE Bernoulli(p) R.V.S

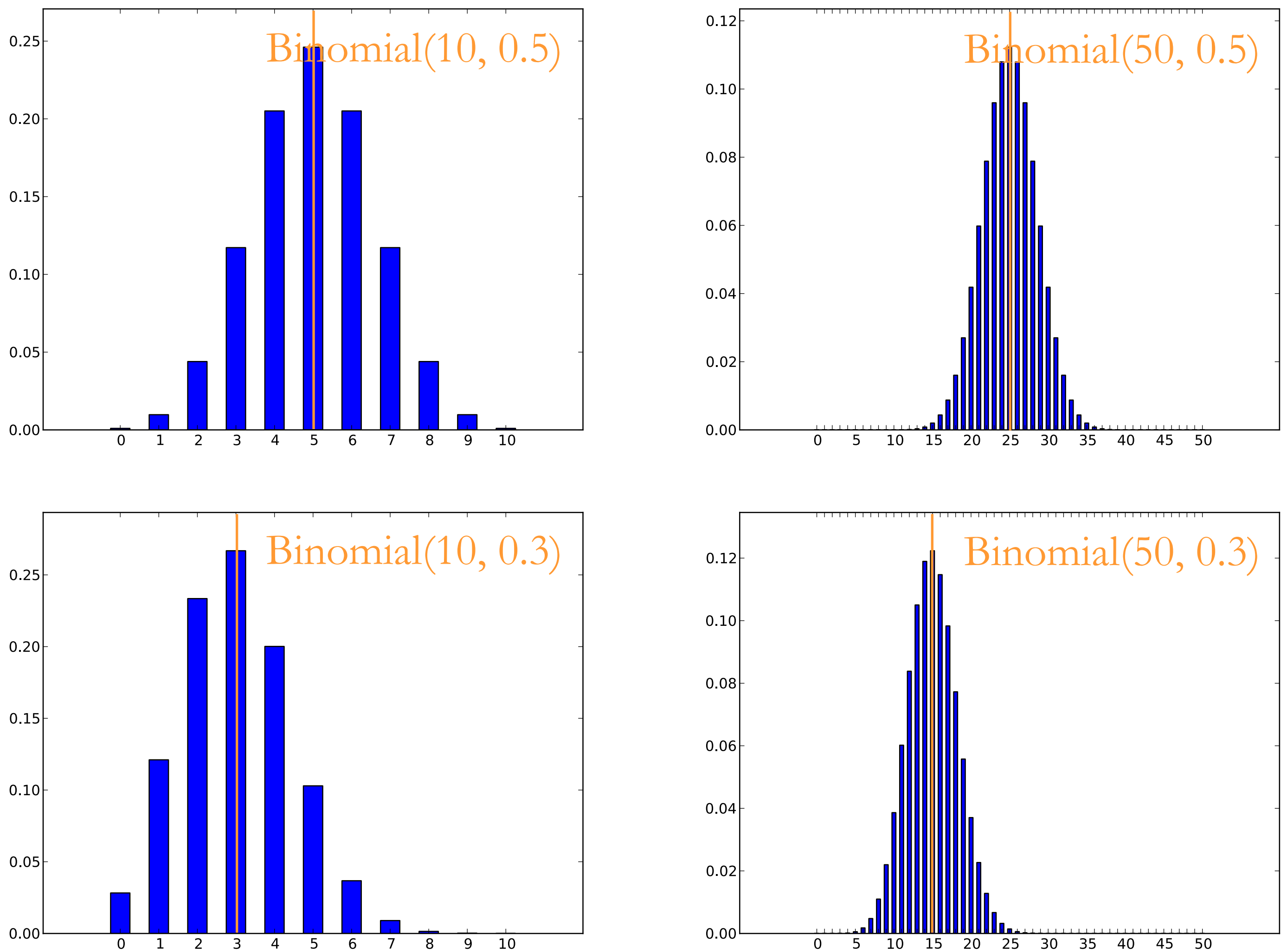
$E[X] =$

$E[X_1] + E[X_2] + \cdots + E[X_n]$

$= p + p + \cdots + p$

$= np$

$$\boxed{E[X] = n \cdot p}$$


Binomial(10, 0.5)
Binomial(50, 0.5)
Binomial(10, 0.3)
Binomial(50, 0.3)
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3 4 5 6 7 8 9 10
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0 5 10 15 20 25 30 35 40 45 50

n

n people throw their hats in a box and pick one out at random. How many on average get back their own?

X = # PEOPLE THAT GET OWN HAT

$X = X_1 + \ldots + X_n$

$X_i = \begin{cases} 1 & \text{IF } i\text{TH PERSON GOT OWN HAT} \\ 0 & \text{IF NOT} \end{cases}$

$E[X] = E[X_1] + E[X_2] + \ldots + E[X_{10}]$

$= \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}$

$= 1$

Mean of the Poisson

Poisson(λ) **approximates** Binomial(n, λ/n) **for large** n

$$\boxed{p(k) = e^{-\lambda} \lambda^k / k!} \qquad k = 0, 1, 2, 3, \ldots$$

$$E[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\lambda} \cdot \lambda^k}{(k-1)!}$$

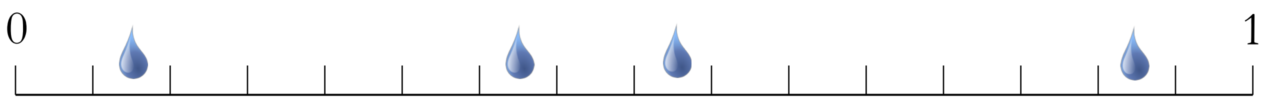
$$= \lambda \cdot \sum_{k=1}^{\infty} \frac{e^{-\lambda} \cdot \lambda^{k-1}}{(k-1)!} = \lambda$$

$$E[\text{Binomial}(n, \lambda/n)] = \lambda$$

$$E[X] = \lim_{n\to\infty} E[\text{Binomial}(n, \lambda/n)]$$

Raindrops

Rain is falling on your head at an average speed of 2.8 drops/second.



X_1 X_2 X_{100}

X_i = PRESENCE OF DROP IN i-TH CENTISECOND

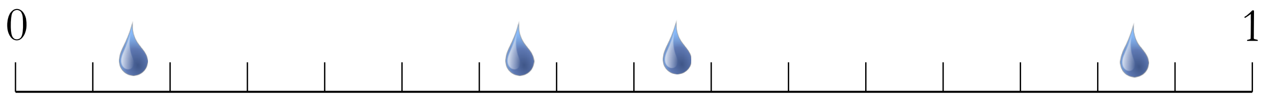
MODEL : $X = X_1 + \ldots + X_{100}$ INDEPENDENT

Binomial$(100, p)$ $100p = 2.8$

$p = 2.8/100$

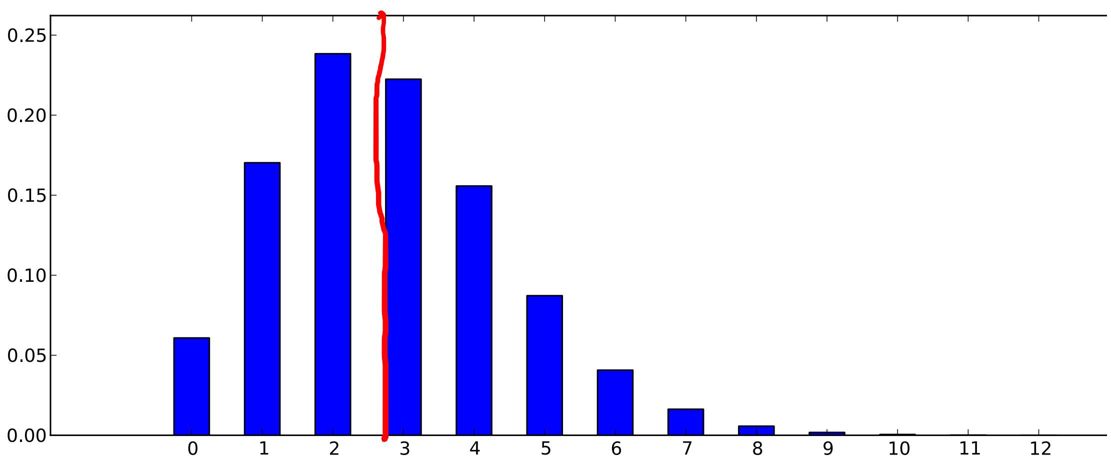
Poisson(2.8): $\lim_{n \to \infty}$ Binomial$(n, 2.8/n)$

Raindrops



Number of drops N **is** $\mathrm{Binomial}(n, 2.8/n)$

$E[X] = 2.8$

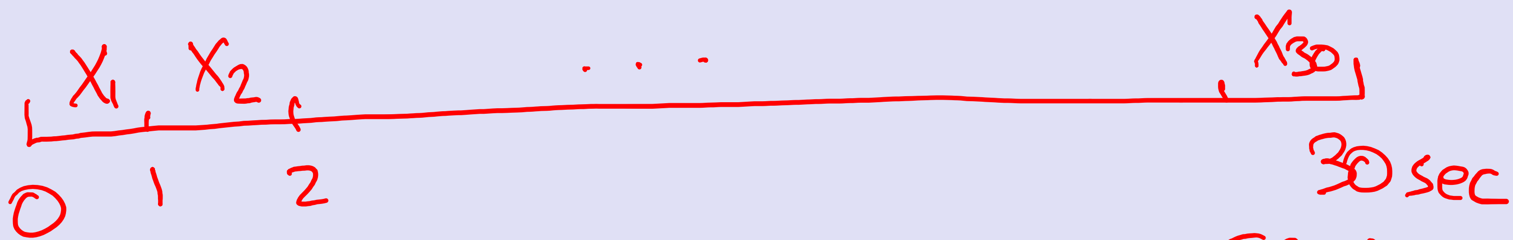


Rain falls on you at an average rate of 3 drops/sec.

When 100 drops hit you, your hair gets wet.

You walk for 30 sec from MTR to bus stop.

What is the probability your hair got wet?



X_i = NUMBER OF DROPS IN i-TH SECOND

$$X = X_1 + \dots + X_{30}$$

$$E[X] = E[X_1] + \dots + E[X_{30}] = 30 \cdot 3 = 90$$

MODEL: X IS Poisson(90)

$$P(X > 100) = 1 - P(X \le 100) = 1 - \sum_{k=0}^{100} e^{-90} \cdot \frac{90^k}{k!}$$

$$\approx 0.134$$

Investments

You have three investment choices:

A: put $25 in one stock

B: put $½ in each of 50 unrelated stocks

C: keep your money in the bank

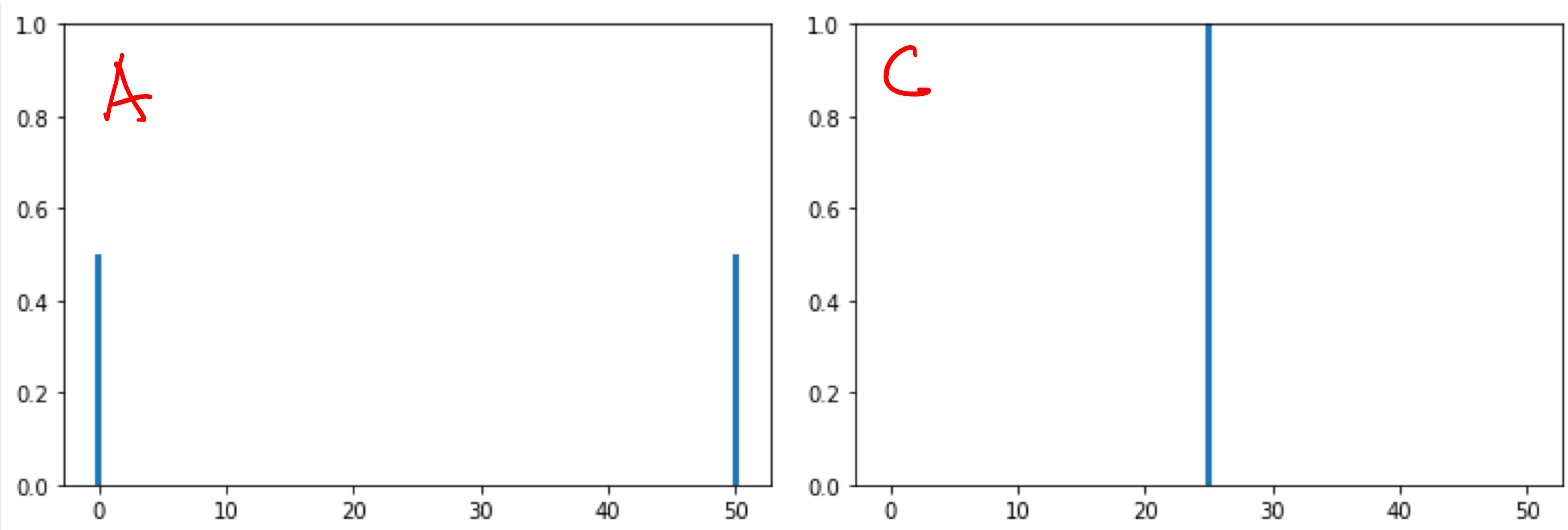
Which do you prefer?

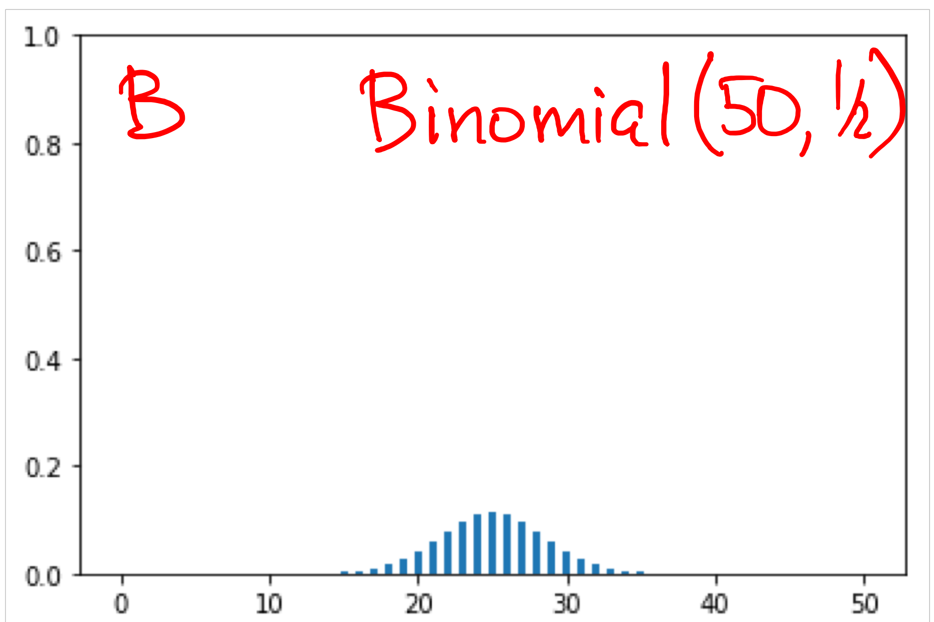
Investments

Probability model

Each stock $\begin{cases} \text{doubles in value with probability } \frac{1}{2} \\ \text{loses all value with probability } \frac{1}{2} \end{cases}$

Different stocks perform independently





$E[A] = 25$

$E[B] = 25$

$E[C] = 25$

Variance and standard deviation

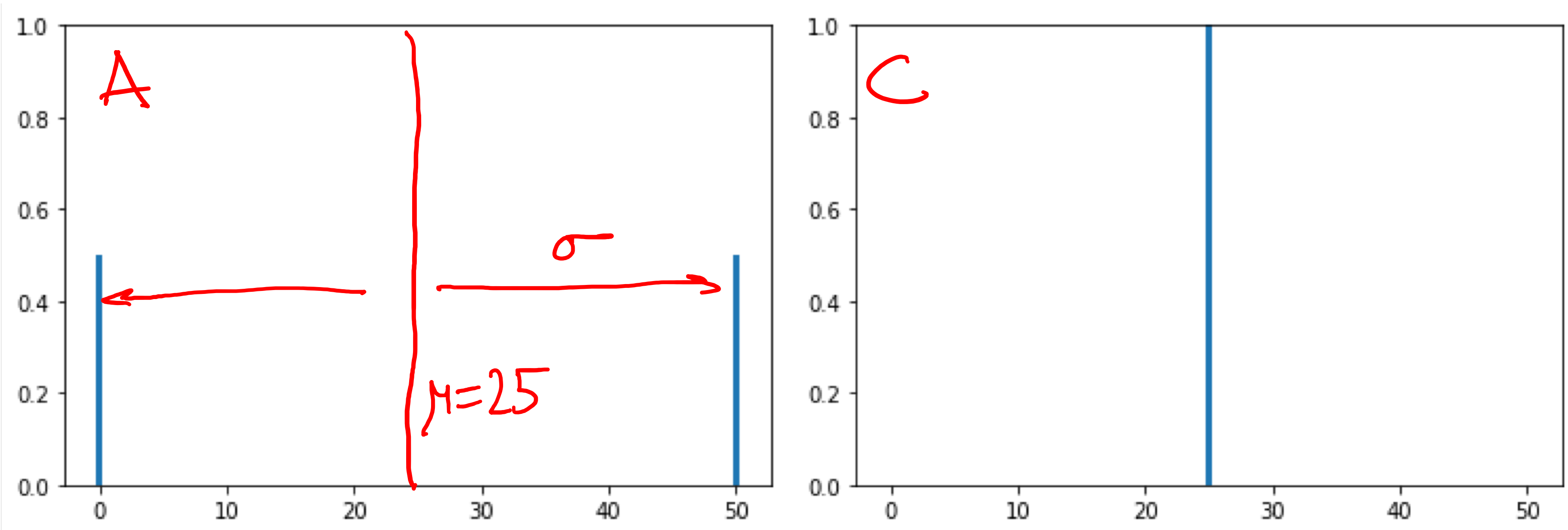
Let $\mu = \mathbf{E}[X]$ be the expected value of X.

The variance of X is the quantity

$$\mathbf{Var}[X] = E[(X - \mu)^2]$$

The standard deviation of X is $\sigma = \sqrt{\mathbf{Var}[X]}$

It measures how close X and μ are typically.



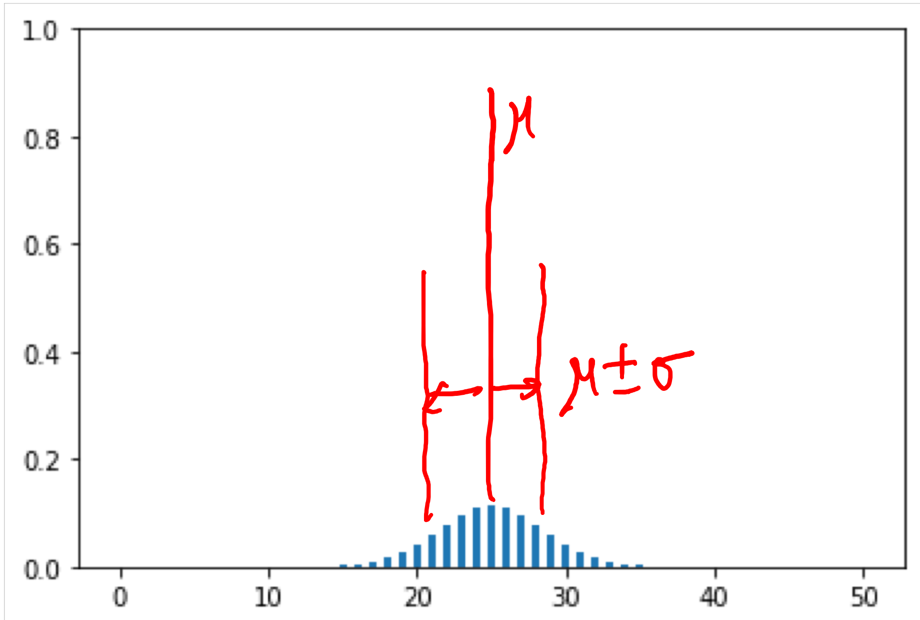
a	0	50
$p(a)$	$1/2$	$1/2$

c	25
$p(c)$	1

$$\mathrm{Var}[A] = \frac{1}{2}(-25)^2 + \frac{1}{2}\cdot 25^2$$
$$= 25^2$$
$$\sigma = 25$$

$$\mathrm{Var}[C] = 0$$

$$\sigma = 0$$



$$\boxed{\mathbf{Var}[\text{Binomial}(n, p)] = np(1-p)}$$

$$\text{Var}[B] = 50 \cdot \frac{1}{2} \cdot \frac{1}{2} = 12.5$$

$$\sigma = \sqrt{12.5} \approx 3.3$$

Another formula for variance

$$\mathrm{Var}[X] = E[(X-\mu)^2] \qquad \mu = E[X]$$

$$= E[X^2 + \mu^2 - 2 \cdot X \cdot \mu]$$

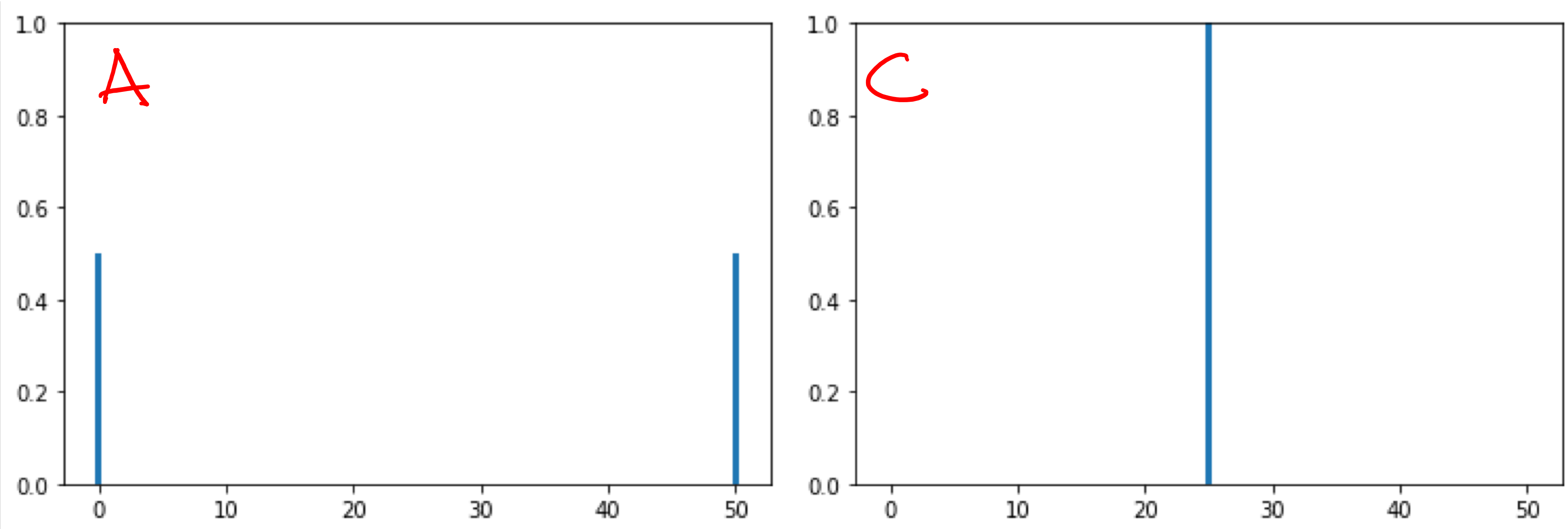
$$= E[X^2] + E[\mu^2] - E[2 \cdot X \cdot \mu]$$

$$= E[X^2] + \mu^2 - 2 \cdot \mu \cdot E[X]$$

$$= E[X^2] + \mu^2 - 2 \cdot \mu \cdot \mu$$

$$= E[X^2] - \mu^2$$

$$= E[X^2] - E[X]^2 \geq 0$$



a	0	50
p(a)	1/2	1/2

$E[A] = 25$

$E[A^2] = \frac{1}{2} \cdot 0^2 + \frac{1}{2} \cdot 50^2$

$Var[A] = \frac{1}{2} \cdot 50^2 - 25^2 = 25^2$

c	25
p(c)	1

$E[C] = 25$

$E[C^2] = 25^2$

$Var[C] = 25^2 - 25^2 = 0$

$$\mathbf{E}[X] = ?$$

$$\mathbf{Var}[X] = ?$$

$$E[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = 3.5$$

$$E[X^2] = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = \frac{1}{6} \cdot \frac{6 \cdot 7 \cdot 13}{6} = \frac{7 \cdot 13}{6}$$

$$Var[X] = \frac{7 \cdot 13}{6} - 3.5^2$$

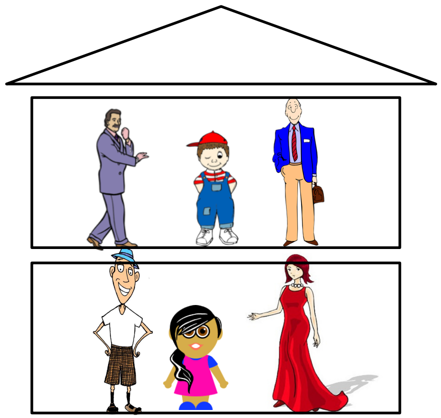
In 2011 the average household in Hong Kong had 2.9 people.

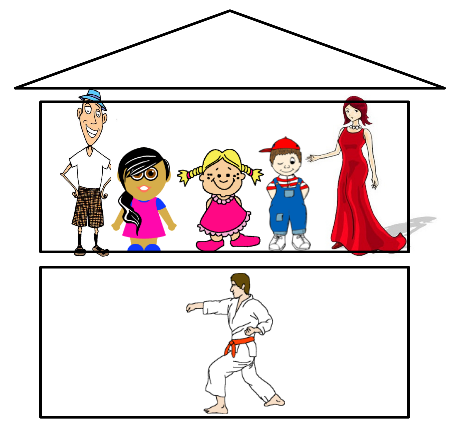
Take a random person. What is the average number of people in his/her household?

A: < 2.9

B: 2.9

C: > 2.9





3

average household size

3

3

average size of random person's household

$$\frac{5}{6}\cdot 5+\frac{1}{6}\cdot 1$$

$$= \frac{26}{6} = 4.33\ldots$$

What is the average household size?

household size	1	2	3	4	5	~~more~~ 6
% of households	16.6	25.6	24.4	21.2	8.7	3.5

From *Hong Kong Annual Digest of Statistics*, 2012

Probability model 1

Households under equally likely outcomes

$X =$ **number of people in the household**

$E[X] = 1 \cdot 0.166 + 2 \cdot 0.256 + \ldots + 5 \cdot 0.087 + 6 \cdot 0.035$

$= 2.903$

What is the average household size?

household size	1	2	3	4	5	more
% of households	16.6	25.6	24.4	21.2	8.7	3.5

y

$P(X=1)$ $P(X=2)$

Probability model 2

People under equally likely outcomes

$Y =$ **number of people in the household**

$$E[Y] = 1 \cdot P(Y=1) + 2 \cdot P(Y=2) + \dots + 6 \cdot P(Y=6)$$

$$P(Y=y) = \frac{\#\text{PPL IN } y\text{-PERSON HOUSES}}{\#\text{PPL}} = \frac{y \cdot P(X=y)}{1 \cdot P(X=1) + \dots + 6 \cdot P(X=6)}$$

$$P(Y=1) = \frac{1 \cdot 16.6}{1 \cdot 16.6 + 2 \cdot 25.6 + \dots + 6 \cdot 3.5}$$

$$E[Y] = \sum_y y \cdot P(Y=y)$$

$$= \sum_y y \cdot \frac{y \cdot P(X=y)}{\sum_x x \cdot P(X=x)}$$

$$= \frac{\sum_y y^2 \cdot P(X=y)}{\sum_x x \cdot P(X=x)}$$

$$= \frac{E[X^2]}{E[X]^2} = \frac{10.213}{2.903} \approx 3.518$$

Summary

$X =$ **number of people in a random household**

$Y =$ **number of people in household of a random person**

$$\mathbf{E}[Y] = \frac{\mathbf{E}[X^2]}{\mathbf{E}[X]} \geq \frac{E[X]^2}{E[X]} = E[X]$$

Because $\mathbf{Var}[X] \geq 0$,

$$\mathbf{E}[X^2] \geq (\mathbf{E}[X])^2$$

So $\mathbf{E}[Y] \geq \mathbf{E}[X]$. The two are equal only if all households have the same size.